

# Unsupervised Well Clustering: Pattern Recognition in Overpressure Mechanisms

Xuan Qin\*, Yan Xu, Huizhong Yan, and De-hua Han, University of Houston

## Summary

In this study, we provide a solution to well clustering by using unsupervised algorithm to recognize patterns in the model-based features. Combining the model-driven feature and data-driven clustering method delivers us robust results of well classification based on overpressure mechanisms. Four types of well are labeled, corresponding to four pore pressure scenarios: normal compaction or disequilibrium compaction, fluid expansion, fluid loss, and a mixture of fluid expansion and loss. The developed procedures assist people to rapidly determine which empirical relations to use while performing pore pressure prediction based on the neighboring well. The clustering results also provide an idea for the lateral variation in a basin.

## Introduction

Overpressure widely exists from the active plate boundaries to passive basins, which has been extensively studied to prevent from geohazards (Swarbrick and Osborne, 1998; Chatterjee et al., 2011) and acquire the status of pore fluids presence, accumulation, and migration (Bruce, 1984; Hao et al., 2007). It tends to happen once the pore fluid pathways to the surface, e.g. sea bottom for offshore or water table for onshore, were cut off so that the interstitial fluid could not escape to restore the hydrostatic pressure.

Overpressure can occur in shales in the process of mechanical compaction, expansion of pore fluids from physical or chemical changes, or both (Swarbrick and Osborne, 1998; Dutta, 2002; Mukerji et al., 2002). Some common mechanisms include disequilibrium compaction or undercompaction, kerogen-to-gas, smectite-to-illite or clay diagenesis, aquathermal expansion, tectonic uplift or erosion, hydrocarbon buoyancy, and vertical or lateral transfer. These mechanisms can have different geophysical signatures (Bowers, 2001; Katahara, 2003; Ramdhan and Goult, 2011; Qin and Han, 2016) and these features may be utilized to establish various empirical relations to predict pore pressure (Eaton, 1975; Dutta, 1986; Bowers, 1995; Lahann et al., 2001). Therefore, to perform reliable pore pressure prediction, we need to understand the mechanisms of overpressure in a region and their effects on physical properties of sedimentary rocks and thus geophysical measurements.

We first select features that can serve as the evaluation criterion for unsupervised learning. Then we extract features from each well based on previous experience or our model-mind. Next, we use the K-means clustering, an unsupervised clustering algorithm, to assess the extracted features and

label all eligible wells into four classes. At last, we show the well data from each cluster to analyze and interpret the data trend within a single well.

## Study Area and Data

Our study area is located in the northern Gulf of Mexico (GoM), where overpressure exists in a wide range and can be caused by two major mechanisms, such as disequilibrium compaction and smectite-to-illite (Verm et al., 1998). Disequilibrium compaction takes place in young sediments with fast sediment rate, and smectite-to-illite (Lahann et al., 2001; Katahara, 2003; Lahann and Swarbrick, 2011, Yu and Hilterman, 2014) happens in smectite-rich shale when formation temperature is above 70 °C. During the smectite-to-illite, an internal increase of pore pressure can be quickly built up within a narrow thickness window (several hundred meters). And the smectite-to-illite transition zone (TRZ) onset depth ranges from 2400 to 3000 m in GoM.

We utilize wireline logging data from 797 wells from the shelf of GoM, offshore Louisiana. The data sets have deleted hydrocarbon zones and defined sandstone as sediments with a shale volume less than 50% and shale with a shale volume larger than 50%. The logging data is averaged on 60 m (200 ft) intervals, including P-wave velocity, resistivity, and density for sand and shale lithology, to filter out high-frequency fluctuations in rock properties that obscure the overall pattern; mud weights, temperature, and overpressure onsets are read from the logging run. With the upscaled data of water-saturated shale formations, it is convenient for us to study the compaction trend in this area.

## Methodology

### Step I: Feature Selection

Analysis with the crossplot of sonic velocity or travel time (DT) against density (Bowers, 2001; Lahann et al., 2001; Hoesni, 2004; Katahara, 2006; Swarbrick, 2012; Tingay et al., 2013; John et al., 2014) is a valid way to discern overpressure mechanism in shale even if pore pressure or effective stress data is not available. With the DT (slowness or inverse of sonic velocity) and density data under a normally pressured condition, a loading curve or normal compaction trend can be defined and fitted as a line in DT-density domain (Katahara, 2006).

If disequilibrium compaction occurs, a mild build-up of overpressure has the effect of slowing the increase rate of velocity and density, while rapid build-up of overpressure can expect reversals of velocity and density. But the data

## Well Clustering for Overpressure Mechanisms

points affected by disequilibrium compaction remains on or near the normal compaction trend (Figure 1a).

Qin and Han (2016) categorized two trends of smectite-to-illite transition with logging data from offshore Louisiana. During the transition, smectite releases interlayer bound water and reacts with potassium ion to form illite. If the released bulk water is preserved in the rock pore system, it can be seen as a process of fluid expansion (Figure 1bFigure 3), which causes more DT increase (a reversal in sonic velocity) and little density change as well as relatively high pressure. If the released water can escape, it can be seen as a process of fluid loss (Figure 1c); density increases and DT changes little while overpressure is relatively small. Solid arrows in Figure 1 denote the end-member cases of fluid expansion and fluid loss. However, there might be data trend in a well corresponding to a case when a part of released water escapes and another part is preserved. So the dashed arrows represent some other possible signatures of velocity and density in a mixed case of fluid expansion and fluid loss. Further loading of transformed clay or illite-rich shale will move along a trend (light blue line in Figure 1) subparallel to the original loading curve (smectite-rich shale, dark blue line in Figure 1).

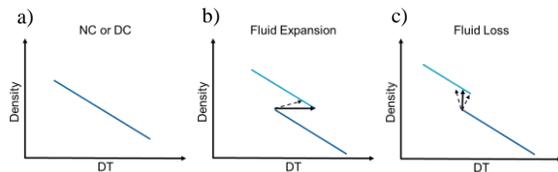


Figure 1. Sketches of data trends in the crossplot of DT (sonic travel time) and density for a) normal compaction (NC) or disequilibrium compaction (DC), b) fluid expansion during smectite-to-illite, and c) fluid loss during smectite-to-illite.

Hence, to cluster well based on their overpressure mechanism, it is significant to grab the data trend shape in entire depth column and data especially in TRZ. It is noteworthy to ask questions include: whether there are two loading curves in shallow and deeper depth, whether there exist a TRZ between the two loading curve and what its data trend.

### Step II: Feature Extraction

With above analysis, we propose to automatically fit linear lines to the trends and extract the incline angles of lines as features. The incline angle represents the direction of a trend. However, because the ranges of DT and density values are different, we rescale DT to density ( $\text{kg/m}^3$ ) with a linear transform, so that the incline angles corresponding to the fitting lines' slopes can range from 0 to  $180^\circ$ . After we rescale DT trend, the vertical trend, e.g. pure fluid loss, has an angle close to  $90^\circ$  in DT-density domain and horizontal trend, e.g. pure fluid expansion, has an incline angle close to  $0^\circ$ . The trends of smectite-to-illite can be modeled by three

sequential lines with the second line lying in the transition zone and the first and third lines characterizing data trend in shallow and deep sections respectively and approximately paralleling with each other. The challenge is to identify the appropriate set of data points that correspond to the transition zone, which can vary from one well to another.

To successfully identify the depths corresponding to the transition zone of each well, we explore different depth ranges for each well, and select the best depth range that leads to the best  $R^2$  fitting score. The onset depth range starts from 2400 to 3000 m with a depth interval varying from 500 to 700 m. To avoid the second line (in the TRZ) deprecated to be aligned with the first line before reaching the transition zone, we first limit the incline angle of second line to the ranges of  $[0^\circ, 115^\circ]$  and  $[165^\circ, 180^\circ]$ , where the angle of first line seldom fall in this range. If all angles of second lines during the test do not fall in the limited range, then limitation is relaxed and second line will have an angle within  $[115^\circ, 165^\circ]$ , nearly aligned with the first line. The flowchart of feature extraction for line 2 is shown in Figure 2. Linear regression is a regression function with DT as X, Density as Y and it returns the incline angle of the fitted linear line as angle and  $R^2$  fitting score as score.

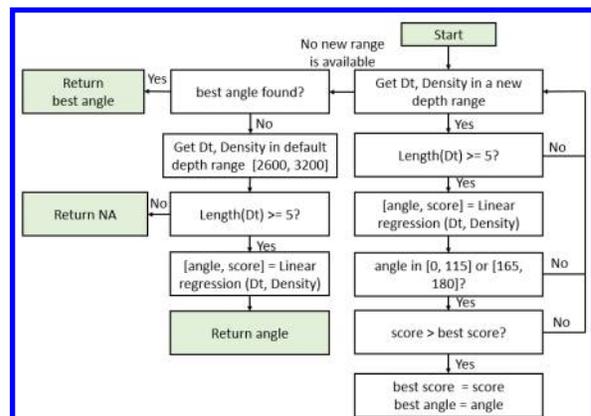


Figure 2. Flow chart for incline angle calculation for the transition zone with varying depth range.

### Step III: Unsupervised Clustering

We propose to use an unsupervised clustering algorithm to help the interpreters in an objective, efficient, and robust way. We use the three inclining angles of the auto-fitted linear lines as input features for each well and use an unsupervised clustering algorithm: K-means (Arthur and Vassilvitskii, 2007) to cluster the wells into similar groups. K-means is a popular clustering algorithm that runs in an iterative way. The number of clusters is pre-determined and the centers of clusters are randomly initialized. At each iteration, each well is assigned to the nearest center based on the three angles and then the center is updated accordingly by taking the average of all assigned wells. Then the wells

## Well Clustering for Overpressure Mechanisms

are reassigned to the updated centers. The iteration continues until the centers do not move or a preset number of iteration is reached.

### Results and Analysis

The well clustering result is shown in Figure 3. The vertical axis is the well index of 90 eligible wells. The horizontal axis represents the incline angles of 3 linear regressions describing data trend in shallow, TRZ, and deep depth of each well. The incline angle is color-coded.

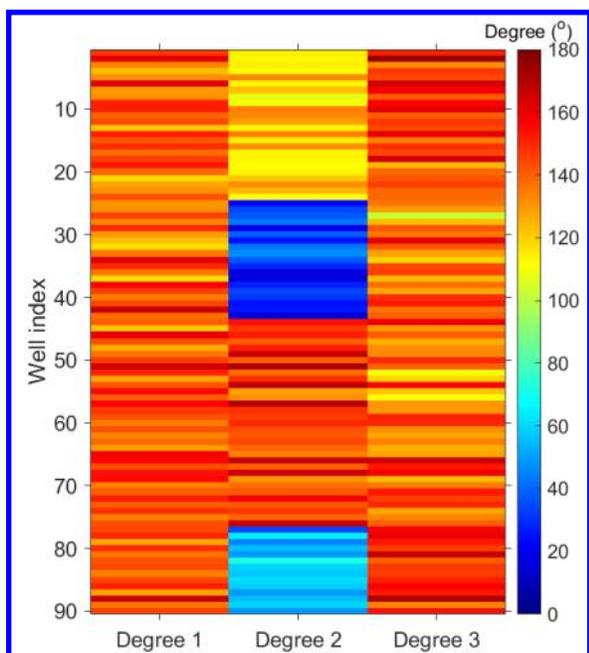


Figure 3. Well clustering result. 90 wells are classified into four groups, Degree 1, 2, and 3 represent the incline angles of 3 linear regressions for DT (rescaled to density range) and density data at shallow, middle, and deep depth range. The middle depth range corresponds to TRZ, which is not fixed, and we seek the best fitting result from a series of TRZ onset depth (from 2400 to 3000 m) and TRZ thickness (from 500 to 700 m).

From the clustering result, we can observe that the incline angles of the first and third linear regression are close and larger than  $110^\circ$ , consistent with the data trend of compaction that DT decrease and density increase. The second linear regression for possible transition zone plays a key role in distinguishing wells to different clusters. Four clusters can be clearly observed in Figure 3. From top to bottom, the first cluster has an incline angle of TRZ smaller than shallow and deep sections but larger than  $90^\circ$ , suggesting fluid loss is the common signature among these wells. Due to the rescaling effect of DT to density, their incline angles are not completely vertical. The second cluster has an incline angle of TRZ close to the horizontal direction,

reflecting fluid expansion is the phenomenon behind this group. The incline angles of TRZ in the third cluster are similar to the incline angles of their shallow and deep sections, indicating disequilibrium compaction or normal compaction occurs in this group. For the last cluster at the bottom, the incline angles of TRZ range from  $50$  to  $90^\circ$ , implying that a mixture of fluid expansion and fluid loss exist in TRZ. Below we present some examples for each cluster.

In Figure 4, we plot two examples of the first cluster labeled with 0. Well log data and three linear regressions for shallow, TRZ, and deep sections are shown in the domain of density and rescaled DT. Depth in meters is color-coded. Each  $R^2$  of linear regressions is also listed. Their TRZs are determined at 2600 to 3100 m and 2400 to 2900 m, respectively. The DT and density data define two subparallel compaction trends at shallow ( $< 2600$  m) and deep (deeper than their transition zone) sections. The incline angles of their TRZs are a little larger than  $90^\circ$  but smaller than their defined compaction trends' incline angles. We interpret these characteristics with the occurrence of smectite-to-illite transition with fluid loss.

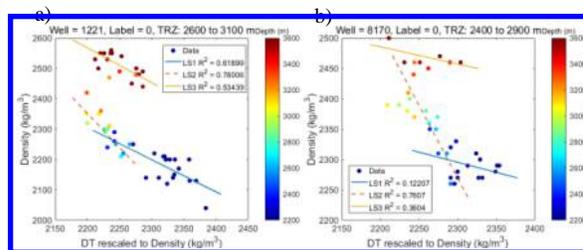


Figure 4. The example wells of the first cluster labeled with 0, corresponding to a major overpressure mechanism as smectite-to-illite fluid loss.

In Figure 5, we plot two examples of the second cluster labeled with 1. Their TRZ are determined at 2600 to 3100 m and 2500 to 3000 m, respectively. The incline angles of their TRZs are smaller than their defined compaction trends' incline angles and a little larger than  $90^\circ$ . We interpret these characteristics with the occurrence of smectite-to-illite transition with expansion.

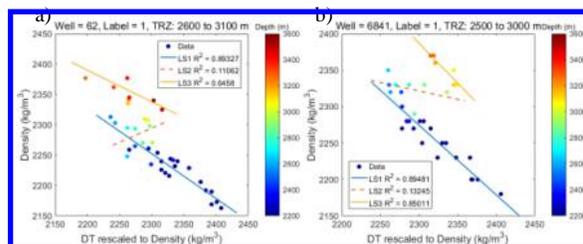


Figure 5. The example wells of the second cluster labeled with 1, corresponding to a major overpressure mechanism as smectite-to-illite fluid expansion.

## Well Clustering for Overpressure Mechanisms

In Figure 6, we plot two examples of the third cluster labeled with 2. Their TRZ are both determined at 2600 to 3200 m. However, the incline angles of their TRZs are almost the same as their compaction trends' incline angles. Therefore, for this group, smectite-to-illite should be not the major mechanism. They have a signature of disequilibrium compaction or normal compaction. Further operations can distinguish whether shales in these wells are in normal compaction or disequilibrium compaction. If shale data is concentrated on the DT-density crossplot with increasing depth, disequilibrium compaction takes place at the concentrated data's depth.

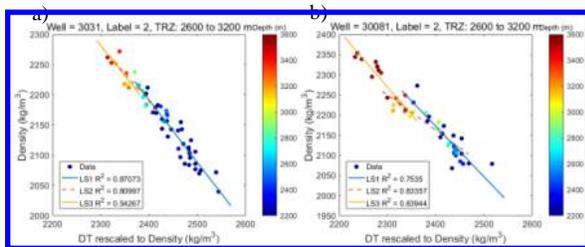


Figure 6. The example wells of the third cluster labeled with 2, corresponding to a major overpressure mechanism as disequilibrium compaction or normal compaction.

In Figure 7, we plot two examples of the fourth cluster labeled with 3. Their TRZ are determined at 2800 to 3400 m and 2800 to 3500 m, respectively. The incline angles of their TRZs are in the range of  $50^\circ$  to  $90^\circ$  and between the incline angles of fluid expansion and fluid loss. We interpret these characteristics as a result of a mixture of fluid expansion and fluid loss.

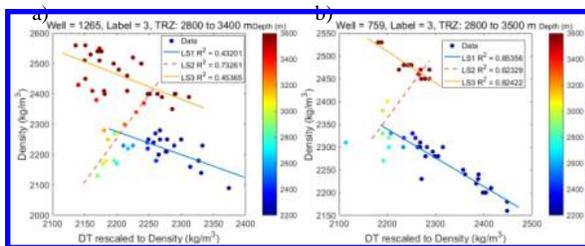


Figure 7. The example wells of the fourth cluster labeled with 3, corresponding to a major overpressure mechanism as a mixture of fluid expansion and loss.

### Discussions

Although the study area contains 2 major overpressure mechanisms, disequilibrium compaction and smectite-to-

illite (Verm et al., 1998; Yu and Hilterman, 2014), 4 clusters of wells can be obtained based on the proposed method. Smectite-to-illite zone have different data trends in TRZ, resulted from fluid expansion, fluid loss, and a mixture of them. This is consistent with Bruce (1984), which suggested that water expelled from smectite during the process of smectite-to-illite might migrate out of the host rock early or might be totally or partially trapped and released gradually during the geological time. Therefore, transferred water distribution during smectite-to-illite might be the first-order factor controlling the rock properties and the concomitant overpressure.

Some technical details can be improved. For instance, the linear regression base on L-2 norm has its limitation to fit the trending line at the assigned depth range, because it gives all the point (data and outliers) the same weight when deriving a model. Moreover, we should be careful when rescaling sonic travel time (DT) to density, because the calculation of incline angles from slopes of fitting lines may affect the clustering result. Last but not the least, data depth is another important parameter, and further work can include depth while extracting features from each well.

Machine learning algorithms perform well in pattern recognition in high-dimensional data, which should be an effective tool in geology and geophysics field. However, machines do not know which feature can be utilized for a specific question. Feature selection with previous experience or a model mind in certain targets facilitates a robust and interpretable learning project.

### Conclusions

Two major overpressure mechanisms, disequilibrium compaction and smectite-to-illite, exists in the study area. Four clusters of wells are concluded based on their DT-density data trends, corresponding to 1) disequilibrium compaction or normal compaction, 2) fluid expansion, 3) fluid loss, and 4) a mixture of fluid expansion and loss within smectite-to-illite transition zone. The data trend within the depth range of possible transition zone (2500 to 3500 m) plays a key role in pattern recognition of overpressure mechanism and well clustering.

### Acknowledgements

The authors would like to thank Fluids/DHI consortium for finial support. We express our gratitude to Dr. Fred Hilterman and Geokinetics for the use of their well-log library.

## EDITED REFERENCES

Note: This reference list is a copyedited version of the reference list submitted by the author. Reference lists for the 2017 SEG Technical Program Expanded Abstracts have been copyedited so that references provided with the online metadata for each paper will achieve a high degree of linking to cited sources that appear on the Web.

## REFERENCES

- Arthur, D., and S. Vassilvitskii, 2007, k-means++: the advantages of careful seeding: Proceedings of the 18th annual ACM-SIAM symposium on Discrete algorithms. Society for Industrial and Applied Mathematics, 1027–1035.
- Bowers, G. L., 1995, Pore pressure estimation from velocity data: Accounting for overpressure mechanisms besides undercompactions, *SPE Drilling & Completion*, **10**, 89–95, <https://dx.doi.org/10.2118/27488-PA>.
- Bowers, G. L., 2001, Determining an appropriate pore-pressure estimation strategy, Offshore Technology Conference, <https://dx.doi.org/10.4043/13042-MS>.
- Bruce, C. H., 1984, Smectite dehydration-its relation to structural development and hydrocarbon accumulation in northern Gulf of Mexico Basin, *AAPG Bulletin*, **68**, 673–683, <https://doi.org/10.1306/ad461363-16f7-11d7-8645000102c1865d>.
- Chatterjee, R., M. Mukhopadhyay, and S. Paul, 2011, Overpressure zone under the Krishna–Godavari offshore basin: Geophysical implications for natural hazard in deeper-water drilling, *Natural Hazards*, **57**, 121–132, <https://dx.doi.org/10.1007/s11069-010-9659-6>.
- Dutta, N. C., 1986, Shale compaction, burial diagenesis, and geopressures: A dynamic model, solution and some results, in J. Burrus, ed., *Thermal modeling in sedimentary basins: Editions Technip*, 149–172.
- Eaton, B. A., 1975, The equation for geopressure prediction from well logs, Presented at the SPE Annual Technical Conference and Exhibition, SPE, <https://dx.doi.org/10.2118/5544-MS>.
- Hao, F., H. Zou, Z. Gong, S. Yang, and Z. Zeng 2007, Hierarchies of overpressure retardation of organic matter maturation: Case studies from petroleum basins in China, *AAPG Bulletin*, **91**, 1467–1498, <https://dx.doi.org/10.1306/05210705161>.
- Kataraha, K., 2003, Analysis of overpressure on the Gulf of Mexico shelf: Offshore Technology Conference, <https://dx.doi.org/10.4043/15293-MS>.
- Katahara, K., 2006, Overpressure and shale properties: Stress unloading or smectite-illite transition?: 76th Annual International Meeting, SEG, Expanded Abstracts, <https://dx.doi.org/10.1190/1.2369809>.
- Hoesni, M. J., 2004, Origins of overpressure in the Malay Basin and its influence on petroleum systems: Ph.D. Thesis, University of Durham.
- Lahann, R. W., D. K. McCarty, and J. C. C. Hsieh, 2001, Influence of clay diagenesis on shale velocities and fluid-pressure: Offshore Technology Conference, <https://dx.doi.org/10.4043/13046-MS>.
- Lahann, R. W., and R. E. Swarbrick, 2011, Overpressure generation by load transfer following shale framework weakening due to smectite diagenesis: *Geofluids*, **11**, 362–375, <https://dx.doi.org/10.1111/gfl.2011.11.issue-4>.
- Mukerji, T., N. C. Dutta, M. Prasad, and J. Dvorkin 2002, Seismic detection and estimation of overpressures part I: The rock physics basis, *CSEG Recorder*, **27**, 34–57.
- Ramdhan, A. M., and N. R. Gouly, 2011, Overpressure and mudrock compaction in the Lower Kutai Basin, Indonesia: a radical reappraisal: *AAPG Bulletin*, **95**, 1725–1744, <https://dx.doi.org/10.1306/02221110094>.
- Swarbrick, R. E., and M. J. Osborne, 1998, Mechanisms that generate abnormal pressures: An overview, in B. E. Law, G. F. Ulmishek, and V. I. Slavin, eds., *Abnormal pressure in hydrocarbon environments: American Association of Petroleum Geologists Memoir 70*, pp. 13–34.

- Qin, X., and D. Han, 2016, Seismic Characters of Pore Pressure Due to Smectite-to-illite Transition: 78th Annual International Conference and Exhibition, EAGE, Extended Abstracts, <https://doi.org/10.3997/2214-4609.201601462>.
- Verm, R., L. Liang, and F. J. Hilterman, 1998, Significance of geopressure in predicting lithology: The Leading Edge, **17**, 227–227, <https://dx.doi.org/10.1190/1.1437952>.
- Yu, H., and F. J. Hilterman, 2014, The effect of pressure on rock properties in the Gulf of Mexico: Comparison between compaction disequilibrium and unloading: Interpretation, **2**, SB1–SB15, <https://dx.doi.org/10.1190/INT-2013-0067.1>.